DOCUMENT RESUME

ED 416 605 EC 306 153

AUTHOR Ysseldyke, James E.; Olsen, Ken
TITLE Putting Alternate Assessments into Practice: What To Measure and Possible Sources of Data. Synthesis Report 28.
INSTITUTION National Association of State Directors of Special Education, Alexandria, VA.; National Center on Educational Outcomes, Minneapolis, MN.; Council of Chief State School Officers, Washington, DC.
SPONS AGENCY Special Education Programs (ED/OSERS), Washington, DC.
PUB DATE 1997-09-00
NOTE 29p.
CONTRACT H159C50004
AVAILABLE FROM National Center on Educational Outcomes, University of Minnesota, 350 Elliott Hall, 75 East River Road, Minneapolis, MN 55455; phone: 612-624-8561; fax: 612-624-0879; World Wide Web: http://www.coled.umn.edu/NCEO ($10).
PUB TYPE Reports - Descriptive (141)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; Data Collection; *Disabilities; *Educational Assessment; Elementary Secondary Education; *Evaluation Methods; State Standards; *Student Evaluation; Testing Programs
IDENTIFIERS *Alternative Assessment; *Testing Accommodations (Disabilities)

ABSTRACT

This report discusses alternate assessments that are to be used in accounting for the performance and progress of students with disabilities who do not participate in typical state assessments. Alternate assessments are data collection procedures used in place of the typical assessment when students cannot take standard forms of assessment. Four information-gathering procedures that might be used in alternate assessments and the application of these procedures to collect data in broader outcome areas are highlighted in the report. Overall, these approaches and those of states currently developing alternate assessments suggest four assumptions that are the foundation of alternate assessment: (1) alternate assessments should focus on authentic skills and on assessing experience in community and other real life environments; (2) alternate assessment should measure integrated skills across domains; (3) if at all possible, alternate assessment systems should use continuous documentation methods; and (4) alternate assessment systems should include as critical criteria the extent to which the system provides the needed supports and adaptations, and trains the student to use them. Four approaches are described that can be used to collect data for alternate assessments of student performance: observation, recollection (via interview or rating scale), record review, and tests. (Contains 43 references.) (CR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 416 605

Synthesis Report 28

Putting Alternate Assessments into Practice: What to Measure and Possible Sources of Data

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

NATIONAL CENTER ON EDUCATIONAL OUTCOMES

In collaboration with:

Council of Chief State School Officers (CCSSO)

National Association of State Directors of Special Education (NASDSE)

EC 306153

BEST COPY AVAILABLE

Synthesis Report 28

Putting Alternate Assessments into Practice: What to Measure and Possible Sources of Data

James E. Ysseldyke
National Center on Educational Outcomes
University of Minnesota,
College of Education and Human Development

•

Ken Olsen
Mid-South Regional Resource Center
University of Kentucky, Human Development Institute

September 1997

NATIONAL CENTER ON EDUCATIONAL OUTCOMES

The Center is supported through a Cooperative Agreement (#H159C50004) with the Division of Innovation and Development, Office of Special Education Programs, U.S. Department of Education. Opinions expressed herein do not necessarily reflect those of the U.S. Department of Education or Offices within it.

NCEO Core Staff

Robert H. Bruininks
Judith L. Elliott
Ron Erickson
Dorene L. Scott
Martha L. Thurlow, Associate Director
James E. Ysseldyke, Director

Additional copies of this document may be ordered for $10.00 from:

National Center on Educational Outcomes
University of Minnesota • 350 Elliott Hall
75 East River Road • Minneapolis, MN 55455
Phone 612/624-8561 • Fax 612/624-0879
http://www.coled.umn.edu/NCEO

Executive Summary

Personnel in most state departments of education are working on the development of alternate assessments that are to be used in accounting for the performance and progress of students with disabilities who do not participate in typical state assessments. The revised IDEA requires that states have alternate assessments in place by the year 2000. Alternate assessments are data collection procedures used in place of the typical assessment when students cannot take standard forms of assessment. Issues that emerge about the content focus of such assessments relate to curriculum relevance; there are several models available that reflect content beyond the academic skills that are the focus of most state assessments. For students with severe and profound disabilities, a broader set of educational outcomes should be assessed. Four information-gathering procedures might be used in alternate assessments; the application of these procedures to collect data in broader outcome areas is highlighted in the report. Overall, these approaches and those of states currently developing alternate assessments suggest four assumptions that are the foundation of alternate assessments:

1. Alternate assessments focus on authentic skills and on assessing experiences in community and other real life environments.
2. Alternate assessments should measure integrated skills across domains.
3. If at all possible, alternate assessment systems should use continuous documentation methods.
4. Alternate assessment systems should include as critical criteria the extent to which the system provides the needed supports and adaptations, and trains the student to use them.

Four approaches are described that can be used to collect data for alternate assessments of student performance:

- Observation
- Recollection (via interview or rating scale)
- Record review
- Tests

These provide a starting point for states to meet the requirement to report, by the year 2000, on the performance of students with disabilities who cannot participate in regular statewide assessments.

Table of Contents

The Challenges of Alternate Assessments 1
Assumptions About Alternate Assessment 2
The "What" of Alternate Assessment 3
The "How" of Alternate Assessment 5
- Observation 5
- Recollection (Via Interview or Rating Scale) 6
- Record Review 9
- Tests 10

Summary of Assessment Methods 12
Matching Content and Methods 13
Some Final Caveats 16
Summary 18
References 19

The Challenges of Alternate Assessments

Personnel in most state departments of education are busy developing frameworks of educational standards, state assessments, and accountability systems (Roeber, Bond, & Braskamp, 1997). They are specifying the knowledge and skills that students will demonstrate, and working to develop ways of assessing the extent to which students achieve those skills. A common challenge across states has been the development of ways to include students with disabilities in state assessment and accountability systems. Personnel at the National Center on Educational Outcomes have repeatedly shown and called attention to the fact that large numbers of students with disabilities are excluded from state assessment and accountability systems (Erickson, Thurlow, & Thor, 1995; Erickson, Thurlow, Thor, & Seyfarth, 1996). It has been argued that when students with disabilities are out of sight in assessment and accountability systems they are out of mind when policy decisions are made and when educational structures and programs are designed. It has been argued (Ysseldyke, Thurlow, McGrew, & Shriner, 1994; Ysseldyke, Thurlow, McGrew, & Vanderwood, 1994) that large numbers of excluded students could participate in state and national assessments, especially if provided with accommodations (e.g., large print, test items read or signed to them, extended time, separate setting, etc.)

The vexing challenge faced, though, is that there is a small group of students (usually students with severe cognitive deficits or multiple disabilities) for whom standard large-scale testing practices and accommodations just do not work. If policy and program decisions are to reflect the needs of all students, states must have aggregate data on the educational progress and accomplishments of students who typically are excluded. The students we are talking about generally are not working toward a regular high school diploma, and their curriculum often includes life skills not typically found in the general curriculum. Traditional assessment and accountability approaches, even with accommodations, simply are value-limited for these students. Alternative approaches are needed to measure the progress of these students toward important educational outcomes. In this report, we describe assumptions that drive alternate assessment considerations and illustrate broad domains in which these procedures make sense. We also define ways to collect information in alternate assessment systems and provide examples and guidelines that illustrate how these procedures can benefit all students with disabilities.

Assumptions About Alternate Assessment

Alternate assessment is a concept that is still emerging. The phrase *alternate assessment* first appears in the recently reauthorized Individuals with Disabilities Education Act as follows (emphasis ours):

> A. IN GENERAL.—Children with disabilities are included in general State and district-wide assessment programs, with appropriate accommodations, where necessary. As appropriate, the State or local educational agency—
>
> (i) develops guidelines for the participation of children with disabilities in **alternate assessments** for those children who cannot participate in State and districtwide assessment programs; and
>
> (ii) develops and, beginning not later than July 1, 2000, conducts those **alternate assessments**.
>
> B. REPORTS.—The State educational agency makes available to the public, and reports to the public with the same frequency and in the same detail as it reports on the assessment of nondisabled children, the following:
>
> (i) The number of children with disabilities participating in regular assessments.
>
> (ii) The number of those children participating in **alternate assessments**.
>
> (iii)(I) The performance of those children on regular assessments (beginning not later than July 1, 1998) and on **alternate assessments** (not later than July 1, 2000), if doing so would be statistically sound and would not result in the disclosure of performance results identifiable to individual children.
>
> (II) Data relating to the performance of children described under subclause (I) shall be disaggregated
>
> (aa) for assessments conducted after July 1, 1998; and
>
> (bb) for assessments conducted before July 1, 1998, if the State is required to disaggregate such data prior to July 1, 1998. [PL 105-17, Section 612 (a)(17)]

From this mandate and the work that is emerging in Kentucky, Maryland and other states, we can make a number of assumptions:

1. An alternate assessment is an assessment that is used in place of the typical assessment. Data are collected via alternate assessment when students cannot take standard forms of assessment (state tests, district exams, etc.) even with accommodations. Therefore, there must be clear criteria and procedures for making decisions about who participates in alternate assessments (e.g., see

Ysseldyke, Olsen, & Thurlow, 1997).

2. Alternate assessments are curriculum-relevant (i.e., they assess what students are learning to know and do); however, the focus of the curriculum for students who participate in an alternate assessment might differ somewhat from the typical curriculum.
3. Performance on alternate assessments will serve as a substitute for information obtained through typical assessments. The results will be aggregated and interpreted in ways designed to ensure accountability and program improvement.
4. Information gained from alternate assessments will serve as an index of student progress toward meeting standards that are held for all students. Therefore, extensive cross-links are essential in regard to curricula and in regard to accountability for all students.

In the sections that follow, we briefly describe the "what" of alternate assessment (content) before going on to describe the "how" (methods) in a little more detail. We then provide examples of matching the content with the methods. Finally, we suggest some parameters for developing a statewide alternate assessment system.

The "What" of Alternate Assessment

For students with severe disabilities, several issues emerge around the "what" of alternate assessment. These issues relate to curriculum relevance. Students with severe disabilities are often in a curriculum that differs in emphasis from the one that is the course of study for other students. Therefore, the typical test, designed to measure the progress and performance of students in a standard curriculum, often will be out of sync with the curriculum in which such students are enrolled (Brown, Branston, Hamre-Neitupski, Pumpian, Certo, & Grunewald, 1979). Statewide tests focus on academic areas. Language arts, mathematics and writing are almost always included, while science and social studies are included nearly as frequently. Yet, stakeholders identified eight domains of essential and desirable outcomes or results when the National Center on Educational Outcomes (NCEO) conducted a national consensus-building process (Vanderwood, Ysseldyke, & Thurlow, 1993). Yet, all of the areas typically assessed in statewide assessments fall within only one of those domains–the outcome domain defined by NCEO as "Academic and Functional Literacy."

Instructional programs for students with disabilities, and especially for students with severe disabilities, tend to focus equal or greater attention on the other educational outcome domains

(e.g. Personal and Social Adjustment, Contribution and Citizenship, Responsibility and Independence, and Physical Health). For most students, acquisition of skills in these functional living domains is assumed to be the result of incidental learning. As Mercer and Mercer (1993) report, however, functional living skills are essential for successful living in modern society; and for some students with learning problems, they must be taught directly and systematically. Otherwise, the students may never acquire them or may learn them through trial and error, which is both costly and time-consuming.

If assessments are to measure what is taught and what is intended to be learned, and if education agencies are to be accountable for all students, alternate assessment must directly address all of the educational outcome domains. In Table 1, we list the five curriculum-related domains of the NCEO outcome model along with four functional living or life-skills frameworks. The curricular areas in these frameworks would be logical candidates for the content of an alternate assessment system.

Table 1. NCEO's Curriculum-Related Outcome Domains and Five Functional Living Frameworks

NCEO's Curriculum-Related Domains	COACH (Giangreco, Clonginger, & Iverson, 1993)	SYRACUSE GUIDE (Ford et al., 1989)	Falvey (1989)	Kokaska & Brolin (1985)	AUEN (Frey, Burke, Jakworth, Lynch, & Sumpter (1996a, 1996b, 1996c, 1996d)
Academic and Functional Literacy	Communication	Self-management and home living	Community skills	Managing family finances	Community Participation and Use
Personal and Social Adjustment	Socialization	Vocational	Domestic skills	Selecting, managing and maintaining a home	Productivity
Contribution and Citizenship	Personal management	Recreation/ Leisure	Recreation skills	Caring for personal needs	Interpersonal relationships
Responsibility and Independence	Leisure/ Recreation	General community functioning	Employment skills	Raising children–family living	Cognitive functioning
Physical Health	Applied academics	Reading and writing	Motor skills	Buying and preparing food	Domestic living
	Home	Money handling	Communication skills	Buying and caring for clothing	
	School	Time management	Functional academic skills	Engaging in civic activities	
	Community		Developing and fostering friendships	Using recreation and leisure	
	Vocational			Getting around in the community	

The "How" of Alternate Assessment

Assessment is a process of collecting data for the purpose of making decisions about students (Salvia & Ysseldyke, 1995). Salvia and Ysseldyke identify 13 kinds of decisions made using assessment information, and they group these into four categories: Prereferral Classroom Decisions, Entitlement Decisions, Post Referral Classroom Decisions, and Accountability/Outcomes Decisions. It is this last set of decisions we are concerned about in this report. Salvia and Ysseldyke (1995) also identified the four kinds of approaches that are used to gather data on students: observation, recollection (via interview or rating scale), record review, and testing. We use this structure to describe the kinds of data that school districts and state departments of education could collect on alternate assessments of student performance.

Observation

Observations can provide highly accurate, detailed, verifiable information about the person being assessed. Data may be collected using systematic or nonsystematic procedures. In systematic observation the observer gathers data on one or more precisely defined behaviors. The frequency, magnitude, or duration of the behavior is recorded, and comparisons are made either to an absolute or normative standard. Nonsystematic observation is informal observation in which the observer watches an individual in his or her environment and takes notes on the behaviors, characteristics, and personal interactions that seem significant. Nonsystematic observation is anecdotal and can be subjective and unreplicable.

What might observational data look like in an alternate assessment program? The data might consist of narrative recordings of student behavior for a specified period of time. They might also be more systematic, involving the observation of behavior and the completion of a checklist. Judgments about data obtained from both systematic and nonsystematic observation could be made using scoring rubrics or rules.

Additional methods that could be used to gather observational data include videotaping and audiotaping. Assessors need to decide whether such taping would be continuous (and for how many hours or days in a row) or snapshot (e.g., every three hours for 10 minutes, or every three days for two minutes).

Observations can be conducted at school, home, or in a community setting depending on the kind of behavior(s) being observed. Teachers, parents, peers who know the student well, or others could conduct them with special training. Observations could be staged, or they could occur in natural environments (e.g., at home, in school, in social situations, at work). That is, students could be asked to do specific things (e.g., walk to the door) or one could observe

student behavior and see whether they engage in specific things. Or, one could introduce a stimulus or challenge and observe how the student responds.

In many instances the data obtained by means of observations are only as good as the observer's knowledge of normal development. Unless there are very clearly defined scoring rubrics, the observer must rely on his or her knowledge of normal development to know whether what is observed differs from standards (either positively or negatively).

Recollection (Via Interview or Rating Scale)

A second major category of methods for collecting data on student performance and progress involves use of interviews, surveys, or rating scales. People familiar with a student can be asked to recall observations and interpretations of behavior and events, and can complete interviews or rating scales based on their recollections.

When interviews or rating scales are used, data may be collected from the student (self-report or self-assessment); from peers; from teachers, therapists or work-study coordinators; from employers; or from family members. Students might be asked how they are doing, or they might be asked about the extent to which they have developed particular skills. The student might write down his or her answer to such questions, or the examiner might record the student's response. The student or other person might complete a checklist or scale. Data also might be collected from peers. Other students might be asked to rate the development or behavior of the student. Peer ratings are especially helpful in rating development in areas like interpersonal communication skills, social behavior, or physical fitness. Most commonly, however, the information source would be a service provider (e.g., teacher, therapist or work-study coordinator) or a family member.

Interviews may be conducted face-to-face, over the telephone or in small groups. Interviews range in structure from casual conversations to highly structured processes in which the interviewer has a predetermined set of questions that are asked in a specific sequence (Salvia & Ysseldyke, 1995). In general, when one wants eventually to aggregate data from interviews of several students, it is best to use a structured interview.

Rating scales can be considered the most formal kind of interview. They enable one to gather data in a structured, sequenced, and standardized way, and facilitate data aggregation. One common kind of rating scale is that which uses a Likert-scale format in which the rater responds to questions or statements by indicating extent of agreement with the statement. A second type of scale requires the rater to indicate the frequency with which specific behaviors occur. A third type involves rating the extent of assistance that must be provided and the

settings in which the behavior is exhibited. For example, the Performance Assessment for Self-Sufficiency (PASS) (American Institutes for Research, 1993) involves both. A teacher or work-study coordinator uses the following scale to rate performance on skills in daily living, personal and social adjustment, employment, and educational areas:

0. Unable to rate
1. Does not or cannot do
2. Does or can do with extensive assistance or supervision
3. Does or can do with some assistance or supervision
4. Does or can do independently.

In addition, the rater indicates the settings (i.e., school, work place, home, other) in which the rater knows the performance. The Assessment of Unique Educational Needs (Frey, Burke, Jakworth, Lynch, & Sumpter, 1996a, 1996b, 1996c, 1996d) is a standards-based approach that looks at functional skills. There are four versions of the scale, each with identical assessment areas, though differing forms and items at each of the four versions. These are shown in Table 1. The Full Independence version is written to address the needs of students with disabilities who are functioning in the normal range of intelligence. The Functional Independence scale is designed to address the educational needs of students with mild mental impairment or those who function as if they have such an impairment, while the Supported Independence Scale is designed to address the educational needs of students with moderate mental impairment who are expected to require ongoing support in adulthood. The Participation Scale is designed to address the educational needs of students with severe or profound mental impairment who are expected to require extensive ongoing support in adulthood. The teacher rates the student's "Consistency of acceptable performance" on a scale ranging from "rarely or never" to "most often." Teachers also indicate the extent to which they are confident of their ratings. Salvia and Ysseldyke (1995) reviewed the most commonly used behavior rating scales in Chapter 26 of their assessment textbook. These scales are listed in Table 2, together with several adaptive behavior scales and other rating scales.

Sometimes we must interview other people and make judgments about student development based on the information they provide us. One helpful way to do so is by using adaptive behavior scales. Scales like the Responsibility and Independence Scale for Adolescents (Salvia, Neisworth, & Schmidt, 1990), Adaptive Behavior Inventory (Brown & Leigh, 1986a), Scales of Independent Behavior-Revised (Bruininks, Woodcock, Weatherman, & Hill, 1996), Checklist of Adaptive Living Skills (Morreau & Bruininks, 1991), and AAMR Adaptive Behavior Scale—School 2 (Nihara, Leland, & Lambert, 1993a) are individually administered scales that are useful sources of items or subtests that can be used to rate and make judgments about student development. A danger in the use of these is identical to the danger for any and all published measures: their content may not match the content of the curriculum.

Table 2. Behavior Checklists Reviewed in Salvia and Ysseldyke, 1995

Scale	Authors	Behaviors Sampled
AAMR Adaptive Behavior Scale—School 2	Nihara, Leland, & Lambert, 1993a, 1993b	Independent and Responsible Functioning, Physical Development, Language Development, Socialization Behaviors, and Personal-Social Responsibility
Adaptive Behavior Inventory	Brown & Leigh, 1986a, 1986b	Self-Care Skills, Communication Skills, Social Skills, Academic Skills, and Occupational Skills.
Attention Deficit Disorders Evaluation Scale-School Version	McCarney, 1989	Inattention Hyperactivity/Impulsivity
Autism Screening Instrument for Educational Planning	Krug, Arick, & Almond, 1993	Sensory Behaviors, Relating, Body and Object Use, Language, Social/Self Help
Behavior Assessment of System for Children	Reynolds & Kamphaus, 1992	Adaptive Behaviors; Adjustment to Teachers, Students and New Situations; Problem Behaviors; Internalizing and Externalizing Behaviors.
The Behavior Evaluation Scale-2	McCarney & Leigh, 1990	Learning/Self Control, Interpersonal/Social, Inappropriate Behavior under Normal Circumstances, Unhappiness/Depression, Physical Symptoms, Fears
Behavior Rating Profile-2	Brown & Hammill, 1990	Emotional, Behavioral, Personal, or Social Adjustment Problems.
Checklist of Adaptive Living Skills	Morreau & Bruininks, 1991	Adaptive Behavior, Self-Care, Personal Independence, Social Functioning, Work Community, and Residential.
Child Behavior Checklist and 1991 Profile for Ages 4-18	Achenbach, 1991a	Participation in Extracurricular Activities, Social Interactions, School Functioning, Internalizing Problems, Externalizing Problems, Social Problems, Thought Problems, Attention Problems, Sex Problems.
Child Behavior Checklist and 1992 Profile for Ages 2-3	Achenbach, 1992	Anxious/Withdrawn Behavior, Aggressive Behavior, Destructive Behavior, Sleep Problems, Somatic Problems.
The Direct Observation Form	Achenbach, 1986	On Task Behaviors, Problem Behaviors (internalizing and externalizing)
Early Childhood Behavior Scale	McCarney, 1992	Academic Progress (performs tasks independently), Social Relationships, Personal Adjustment
Independence Scale for Adolescents	Salvia, Neisworth, & Schmidt, 1990	Self-Management, Independence, Self-care, Career Skills, and Living Independently
Performance Assessment for Self-Sufficiency	American Institutes for Research, 1993	Daily Living, Personal and Social Development, Employment, Educational Performance, Major Problem Behaviors
Scales of Independent Behavior-Revised	Bruininks, Woodcock, Weatherman, & Hill, 1996	Fine and Gross Motor skills, Social Interaction, Language Comprehension and Expression, Personal Living Skills, Self-Care Skills, Community Living Skills
Systematic Screening for Behavior Disorders	Walker & Severson, 1992	Internal and External Problem Behaviors
Teacher's Report Form and 1991 Profile for Ages 5-18	Achenbach, 1991b	Academic Performance, Adaptive Characteristics, Problem Behaviors
Youth Self-Report and 1991 Profile for Ages 11-18	Achenbach, 1991c	Competence in Extracurricular Activities, Social Competence, Internalizing Behaviors, Externalizing Behaviors

Record Review

A third source of data is existing information. There are five kinds of existing information: school cumulative records, school databases, student products, anecdotal records and non-school records. Use of these data sources for an alternate assessment system requires development of standardized record extraction forms and procedures in order to ensure consistency and utility of the information.

Cumulative records on students with disabilities or separate IEP files include, in addition to standard information, copies of their IEPs and indications of the extent to which they are making progress toward accomplishment of IEP objectives. They also include individualized test scores, multidisciplinary team evaluations, and information about student development. In some cases, a student database might be available for post-hoc analysis (e.g., if student information on goal attainment is kept for tracking and reporting purposes).

A number of attempts have been made to aggregate data on IEPs. These efforts have usually failed for three reasons. First, IEPs vary considerably in specificity. IEPs for one teacher, school or district might be written at a detailed task level while other teachers, schools or districts might write their IEPs at a more general level. Second, IEPs have not typically addressed a student's entire educational experience. The IEP usually focuses only on the aspects of a student's education that require specialized supports and services. Therefore, such an IEP would not allow accountability for a student's progress in areas where the student is not receiving special education. Finally, IEPs are usually developed on an idiosyncratic basis from individual assessments rather than from a common framework or curriculum. Therefore, there is no basis for aggregation. Having said that, we are closely watching developments of a study in Iowa (Grimes, 1996). In this study Grimes reports success in being able to aggregate IEPs for the purposes of statewide accountability.

Besides cumulative records, student products might be a source of data. Students produce many permanent products: drawings, worksheets, writing samples, etc. Some of these products usually are retained by teachers and, especially in the case of multiple products of a similar nature, over time can be used to judge change. Such products are increasingly accumulated into a portfolio that can be used to judge progress. Portfolios are discussed in more detail later in this document.

Finally, most teachers and therapists working with students who have moderate to severe disabilities keep extensive anecdotal records about student performance, behavior, and physical status. With a little more work, information can be obtained from non-school sources—parents, medical personnel and others. This information can be of use in making decisions

about the extent to which students are meeting or making progress toward meeting some standards.

When one relies on records to gather information about student achievement, there are a number of limitations. First, one usually must go through a considerable volume of information in order to gather the data necessary to answer assessment questions. The process takes a considerable amount of time. Second, the assessor has no control over data collected in the past. The person who recorded information has decided what is relevant to record. Third, context formation is critical, but usually impossible to evaluate. It is necessary to know the conditions under which a student demonstrated a behavior or performed a task, yet contextual information typically is not included in student records.

Tests

The final method for gathering achievement information is the most common for most students: testing. Testing is the process of measuring student competencies, attitudes, and behaviors by presenting a challenge or problem and having the student generate a response. Many states now use either norm-referenced tests or performance-based measures to assess student progress toward the attainment of standards (see Roeber et al.) for more information on types of large scale tests). In general, the kinds of tests used by states do not function well for students with more severe disabilities due to the complexities of the tasks, the cognitive skills involved and the content addressed by the tests.

It might be possible to take the tests designed to measure standards and use those tests to gather information on beginning components of the standards. For example, Gerald Tindal, working with personnel in the Oregon Department of Education, suggests that if a performance assessment involves comprehension of written text, a more basic version of that measure might involve reading a passage to a youngster and asking him or her to "Tell me about the story." Based on an explicit set of criteria, the examiner could record information that indicates the extent to which the student understood the story (e.g., by recording the number of relevant words, connected phrases, etc.). Suppose, for example, that the passage being read is "The Three Bears." Relevant utterances might be words like "bear," whereas words like "truck" would be considered incorrect. For the purposes of a statewide alternate assessment, the challenge/problem statements, the criteria and recording techniques would have to be standardized.

A second option might be the use of existing standardized measures. There are no standardized tests that address all five NCEO domains while being appropriate for students at multiple age levels. However, some individual and group measures exist that assess some of the domains

for some age levels. A battery of tests might be selected to collectively assess the content areas.

Increasingly, portfolio systems are being used as tests of student performance and progress. Portfolios might consist partially of tests and partly of naturally occurring records. A number of different models of portfolio assessment have been advocated, and there is little consensus on what constitutes a portfolio or how portfolios should be used in large-scale assessment (Salvia & Ysseldyke, in press; Wolf & Baron, 1996). In Kentucky, student entries in the alternate portfolio vary, but must include a schedule showing the extent to which the student is involved in independent and integrated activities, letters from the family or the caregiver and the students, and at least six other entries. The portfolios are rated based on the extent to which natural supports are accessed, the settings in which the performance is exhibited, the level of interactions with peers without disabilities, the range of contexts used, and the extent of coverage of the state's academic expectations for all students. Regardless of whether the state or local agency chooses to adapt the existing state test, select a battery of published measures, use performance events, or use a portfolio system, a number of testing development and interpretation considerations must be taken into account.

Tests can result in two kinds of information, quantitative and qualitative. For the purposes of this report, quantitative data are the actual scores that students earn, while qualitative data consist of other observations made while a student is tested. Developers must decide whether the qualitative information will be collected and used systematically. For example, the observational data during testing can tell us how the student achieved a particular score (Salvia & Ysseldyke, in press) and such data can be included in a scoring rubric.

Also, the state or local education agency must decide whether to use absolute standards or normative standards in interpreting student performance. In normative assessment, the performance of the individual is compared to the performance of peers. In most cases, states will need to develop their own norms for the population taking the alternate assessment. This will be difficult due to the extreme variability in the population. When absolute standards are used (as in criterion-referenced or curriculum-based assessments), comparison is made to absolute levels of performance. For example, Kentucky and Maryland have developed four-level rubrics for portfolios and, in Maryland, for performance events. For each level (e.g., novice, apprentice, proficient, distinguished), they have identified samples that serve as "benchmarks" or standards against which the performance of all students is judged. Absolute standards also might be implied in the curricular objective, e.g., "Student correctly identifies gender restroom signs in a community setting 100% of the time whether the signs are presented in text or as icons."

Finally, developers must decide how problems will be presented and how responses will be solicited and recorded. Students who are in an alternate assessment often face significant challenges in cognition and communication. Paper and pencil measures are usually inappropriate without use of a scribe. Oral or communication board responses might be required. For students who have extremely limited communication, computer-assisted choice systems might be necessary.

Summary of Assessment Methods

Table 3 is a summary of the various options within the four information-gathering methods. The pros and cons of each method and each option are not presented in this table because they are related to the issue of curriculum relevance.

Table 3. Summary of Assessment Methods

Observations—Teachers or third party informant watching student exhibit the behavior
- Staged or natural
- Taped or live
- Segmented or continuous

Interviews/Surveys—Gathering information by interviews or surveys with people who know the student (caregiver, parent, student, teacher, therapist, work-study coordinator, employer)
- Face-to-face or phone interviews (group or individual)
- Mail surveys
- Standard checklists, rating scales, adaptive behavior records

Record Reviews—Using a structured procedure to extract information
- Cumulative file/IEPs
- Databases
- Student Products
- Teacher/Therapist Anecdotal Records
- Non-school records, e.g., parents' files and medical records

Tests—Putting a challenge in front of students and having them solve the problem
- Adaptations of the state assessment
- Battery of published instruments
- Performance events
- Portfolios
- Close- or open-ended
- Norm or criterion referenced
- Variety of options for communicating responses

Matching Content and Methods

Figure 1 shows a matrix that intersects the five NCEO outcome domains with the four assessment methods. How might a state or local agency apply these four methods to the five NCEO outcome domains as portrayed in Figure 1?

Figure 1. Options for Alternate Assessment

	Observe	Interview or Survey	Review Records	Test
Academic and Functional Literacy				
Personal and Social Adjustment				
Contribution and Citizenship				
Responsibility and Independence				
Physical Health				

In the spring of 1997, personnel from the Mid-South Regional Resource Center addressed that issue. It convened teachers of students with moderate, severe and profound disabilities from five states to generate ways that the skills and knowledge of students who need an alternate assessment might be assessed. The teachers were presented the five content domains from the NCEO model and were asked to generate ideas for using each data collection technique for assessing each domain. The teachers were asked to generate techniques that:

- were appropriate to students with severe disabilities,
- would be feasible to administer on a large scale, and
- were specific to a cell of the matrix in Figure 1 (even though a state system would most likely combine both content areas and methods).

Their ideas, presented in Figures 2 through 5, illustrate the range of options available to a state. For example (see Figure 2), they suggested that if you wanted to use observation to assess Contribution and Citizenship, students could be videotaped participating in several community activities and their involvement could be rated according to some specific criteria.

Figure 2. Example of an Alternate Assessment Using Observation to Assess Contribution and Citizenship

	Observe	Interview or Survey	Review Records	Test
Academic and Functional Literacy				
Personal and Social Adjustment				
Contribution and Citizenship	■			
Responsibility and Independence				
Physical Health				

Videotape each student participating in multiple community activities (e.g., service projects, scouts, 4-H, group nursing home visits) and rate the extent to which the student follows rules, contributes to the group and performs assigned roles.

If you wanted to assess Academic and Functional Literacy via an interview or survey (see Figure 3), people who are directly and regularly involved with the students could be interviewed or could independently complete a checklist about each student's functional skills.

Figure 3. Example of an Alternate Assessment Using a Survey to Assess Academic and Functional Literacy

	Observe	Interview or Survey	Review Records	Test
Academic and Functional Literacy		■		
Personal and Social Adjustment				
Contribution and Citizenship				
Responsibility and Independence				
Physical Health				

Have parents, teachers and therapists complete checklists and rating scales regarding specific functional math skills, use of vocabulary, and basic science skills, judging the student's performance in various settings (e.g., home, school, community).

The teachers suggested that using multiple data sources that already exist might be a way to gather information about a student's current status and progress in the area of Responsibility and Independence (see Figure 4).

Figure 4. Example of an Alternate Assessment Using Record Reviews to Assess Responsibility and Independence

	Observe	Interview or Survey	Review Records	Test
Academic and Functional Literacy				
Personal and Social Adjustment				
Contribution and Citizenship				
Responsibility and Independence			■	
Physical Health				

Review student files and extract data on current status and changes in self-care skills based on IEPs, anecdotal notes, task analysis charts, therapist reports, parent notes/reports, and conference summaries.

If you wished to test Personal and Social Adjustment, the teachers suggested a performance task. A student could be given an errand that required the student to interact with some people with whom the student was unfamiliar. To avoid having to follow the student around, those people would be asked to rate the quality of the interactions and the extent to which the student had available the supports he or she needed to function appropriately (see Figure 5).

BEST COPY AVAILABLE

Figure 5. Example of an Alternate Assessment Using a Performance Test to Assess Personal and Social Adjustment

	Observe	Interview or Survey	Review Records	Test
Academic and Functional Literacy				
Personal and Social Adjustment				■
Contribution and Citizenship				
Responsibility and Independence				
Physical Health				

Assign students a task (e.g., an errand) requiring interaction with persons unfamiliar to them but who are prepared to judge the quality of the interactions and the extent to which the student had the needed supports and accommodations to enable the interactions (e.g., appropriate communication devices).

Some Final Caveats

Gathering data on the performance of students with disabilities through alternate assessments requires some re-thinking of traditional assessment methods. An alternate assessment system is neither a traditional large scale assessment system nor an individualized assessment. Alternate assessments are a highbred—a common assessment that can be administered to students who have a unique array of educational goals and experiences and who differ greatly in their abilities to respond to stimuli, solve problems, and provide responses.

Although the efforts represent different state perspectives, the work of the alternate assessment system developers in Kentucky (Kleinert, Kearns, & Kennedy, in press) and Maryland (Haigh, 1996) and the work of the Mid-South RRC teacher cadre make it apparent that a common set of assumptions or caveats is emerging about the development of these systems.

1. **Focus on authentic skills and on assessing experiences in community/real life environments.**
   The focus of the assessment must be on real life community-based experiences. If students are going to be expected to function in a community, they must be able to perform in real or authentic community situations. Artificial assessment tasks will not provide an indication of how well the system is preparing the

students; however, "community" means different things at primary, middle and secondary levels. For a third grader, community might be the school, the playground and home, whereas community for an exiting senior would have to mean the store, bank, and workplace, for example.

2. **Measure integrated skills across domains.**
   The examples above are not realistic ways to assess these students because education, especially for students with moderate to severe cognitive disabilities, requires integration of skills. So should the assessments. For example, assessing personal and social skills separately from assessing independence and responsibility would result in redundant effort and possibly result in reinforcing a focus on isolated skills. A generic rubric that encompasses multiple skills would be more appropriate.
3. **Use continuous documentation methods if at all possible.**
   Using assessment methods that involve multiple measures over time will result in more accurate and reliable information. Students with severe challenges have greater variability in their skills from day to day than do students without disabilities or even students with milder disabilities. Therefore, a skill that cannot be observed on one day might be fully in place the next. Also, longitudinal data-gathering methods will be more sensitive than snapshot approaches. Milestones for students with severe disabilities are much farther apart than for other students, and methods that capture change rather than status will better reflect success of the educational system.
4. **Include, as critical criteria, the extent to which the system provides the needed supports and adaptations and trains the student to use them.**
   If the purpose is to hold the educational system accountable, the only way to assess the extent to which a school system is providing the needed education is to include, as one of the criteria for success, the extent to which the school system provides the needed assistive devices, people and other supports to allow the students to function as independently as possible. There is more variability in the skill levels and needs of this one percent of the students than there is in the rest of the total student population. Adding an accommodation/ support criterion helps level the playing field so that the most severely involved students do not always receive the lowest scores. Kentucky has shown that including this criterion has the added benefit of driving effective school and classroom practice (Kleinert et al., in press).

Summary

The topic of alternate assessment is on the front burner, fueled by the needs of SEA and LEA personnel to account for the performance and progress of ALL students, including all students with disabilities. The need is exacerbated by the fact that it is now law. By July 1, 2000, states must conduct alternate assessments and report on the results of those assessments.

In this report, we have defined alternate assessments, described a conceptual framework for thinking about them, and provided initial thinking about ways in which data might be collected on educational results for students with severe disabilities. We provide a starting point for personnel in state and local education agencies. We recognize that our thoughts will have to be adapted to meet specific state and local needs.

References

Achenbach, T.M. (1986). The Direct Observation Form (DOF). Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T.M. (1991a). Integrative guide to the 1991 CBCL, YSR, and TRF profiles. Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T.M. (1991b). Manual for the Child Behavior Checklist/4-18. Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T.M. (1991c). Teacher's report form (TRF). Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T.M. (1992). Child behavior checklist/2-3 years (CBCL/2-3). Burlington, VT: University of Vermont Department of Psychiatry.

American Institutes for Research. (1993). Performance assessment for self-sufficiency (PASS). Palo Alto, CA: American Institutes for Research.

Brown, L., Branston, M.B., Hamre-Neitupski, S., Pumpian, I., Certo, N., & Grunewald, L. (1979). A strategy for developing chronological age appropriate and functional curriculum content for severely handicapped adolescents and adults. Journal of Special Education, 13, 81-90.

Brown, L. & Hammill, D. (1990). Behavior rating profile (2nd ed.). Austin, TX: Pro-Ed.

Brown, L. & Leigh, J. (1986a). Adaptive Behavior Inventory. Austin, TX: Pro-Ed.

Brown, L. & Leigh, J. (1986b). The Adaptive Behavior Inventory manual. Austin, TX: Pro-Ed.

Bruininks, R.H., Woodcock, R., Weatherman, R., & Hill, B. (1996). Scales of Independent Behavior-Revised. Chicago, IL: Riverside.

Erickson, R.N., Thurlow, M.L., & Thor, K.A. (1995). 1994 state special education outcomes. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Erickson, R.N., Thurlow, M.L., Thor, K., & Seyfarth, A. (1996). 1995 state special education outcomes. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Falvey, M. (1989). Community-based curriculum: Instructional strategies for students with severe handicaps (2nd ed.). Baltimore: Brookes Publishing.

Ford, A., Schnorr, R., Meyer, L., Davern, L., Black, J., & Dempsey, P. (1989). The Syracuse community-referenced curriculum guide. Baltimore: Brookes Publishing.

Frey, W., Burke, D., Jakworth, P., Lynch, L., & Sumpter, M.L. (1996a). Addressing unique educational needs of individuals with disabilities: Educational performance expectations for achieving full independence in major life roles: AUEN 3.0. Lansing, MI: Disability Research Systems, Inc.

Frey, W., Burke, D., Jakworth, P., Lynch, L., & Sumpter, M.L. (1996b). Addressing unique educational needs of individuals with disabilities: Educational performance expectations for achieving functional independence in major life roles: AUEN 3.0. Lansing, MI: Disability Research Systems, Inc.

Frey, W., Burke, D., Jakworth, P., Lynch, L., & Sumpter, M.L. (1996c). Addressing unique educational needs of individuals with disabilities: Educational performance expectations for achieving participation in major life roles: AUEN 3.0. Lansing, MI: Disability Research Systems, Inc.

Frey, W., Burke, D., Jakworth, P., Lynch, L., & Sumpter, M.L. (1996d). Addressing unique educational needs of individuals with disabilities: Educational performance expectations for achieving supported independence in major life roles: AUEN 3.0. Lansing, MI: Disability Research Systems, Inc.

Giangreco, M., Clonginger, C., & Iverson, V. (1993). Choosing options and accommodations for children: A guide to planning inclusive education. Baltimore: Brookes Publishing.

Grimes, J. (1996). Iowa-special education effectiveness results (I-SEE results), year 1 pilot project, summary report. Des Moines: Iowa Department of Education.

Haigh, J. (1996). Maryland's pilot alternate assessment checklist. Baltimore: Maryland State Department of Education.

Kleinert, H., Kearns, J.F., & Kennedy, S. (In press). Accountability for all students: Kentucky's Alternate Portfolio Assessment for students with moderate and severe disabilities. Journal of the Association for Severe Handicaps.

Kokaska, C.J. & Brolin, D.E. (1985). Career education for handicapped individuals (Second edition). New York: Merrill/Macmillan.

Krug, D.A., Arick, J.R., & Almond, P.A. (1993). Autism screening instrument for educational planning (2nd ed.). Austin, TX: Pro-Ed.

McCarney, S.B. (1989). Attention deficit disorder evaluation scale-school version. Columbia, MO: Hawthorne Educational Services.

McCarney, S.B. (1992). Early childhood behavior scale: Technical manual. Columbia, MO: Hawthorne Educational Services.

McCarney, S.B. & Leigh, J.E. (1990). Behavior Evaluation Scale-2. Columbia, MO: Hawthorne Educational Services.

Mercer, C.D. & Mercer, A.R. (1993). Teaching students with learning problems (4th ed.). Englewood Cliffs, NJ: Merrill.

Morreau, L.E. & Bruininks, R.H. (1991). Checklist of adaptive living skills (CALS) manual. Allen, TX: DLM.

Nihira, K., Leland, H., & Lambert, N. (1993a). AAMR adaptive behavior scale-school (2nd ed.). Austin, TX: Pro-Ed.

Nihira, K., Leland, H., & Lambert, N. (1993b). Examiner's manual, AAMR adaptive behavior scale-residential and community (Second edition). Austin, TX: Pro-Ed.

Reynolds, C. & Kamphaus, R. (1992). Behavior assessment system for children. Circle Pines, MN: American Guidance Service.

Roeber, E., Bond, L., & Braskamp, D. (1997). Annual survey of state student assessment programs. Washington, DC: Council of Chief State School Officers.

Salvia, J. & Ysseldyke, J.E. (1995). Assessment (6th ed.). Boston: Houghton Mifflin.

Salvia, J. & Ysseldyke, J.E. (In press). Assessment (7th ed.). Boston: Houghton Mifflin.

Salvia, J., Neisworth, J., & Schmidt, M. (1990). Examiner's manual: Responsibility and independence scale for adolescents. Allen, TX: DLM.

Vanderwood, M., Ysseldyke, J.E., & Thurlow, M.L. (1993). Consensus building: A process for selecting educational outcomes and indicators (Outcomes and Indicators No. 2). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Walker, H.M. & Severson, H.H. (1992). Systematic screening for behavior disorders (Second edition). Longmont, CO: Sopris West.

Wolf, D. & Baron, J.R. (1996). Performance-based student assessment: Challenges and possibilities. Chicago: University of Chicago Press.

Ysseldyke, J.E., Olsen, K., & Thurlow, M.L. (1997). Issues and considerations in alternate assessments (Synthesis Report 27). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Ysseldyke, J.E., Thurlow, M.L., McGrew, K.S. & Shriner, J.G. (1994). Recommendations for making decisions about the participation of students with disabilities in statewide assessment programs (Synthesis Report 15). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Ysseldyke, J.E., Thurlow, M.L., McGrew, K.S., & Vanderwood, M. (1994). Making decisions about the inclusion of students with disabilities in large-scale assessments (Synthesis Report 13). Minneapolis: University of Minnesota, National Center on Educational Outcomes.

The College of Education
& Human Development

UNIVERSITY OF MINNESOTA

BEST COPY AVAILABLE